Structure, Vol. 13, 1755–1763, December, 2005, ©2005 Elsevier Ltd All rights reserved. DOI 10.1016/j.str.2005.08.009

Slow Protein Conformational Dynamics from Multiple Experimental Structures: The Helix/Sheet Transition of Arc Repressor

Robert B. Best, Yng-Gwei Chen, and Gerhard Hummer[1,*]
Laboratory of Chemical Physics
Building 5
National Institute of Diabetes and Digestive and Kidney Diseases
National Institutes of Health
Bethesda, Maryland 20892

Summary

Conformational transitions underlie the function of many biomolecular systems. Resolving intermediate structural changes, however, is challenging for both experiments and all-atom simulations because the duration of transitions is short relative to the lifetime of the stable species. Simplified descriptions based on a single experimental structure, such as elastic network models or Gō models, are not immediately applicable. Here, we develop a general method that combines multiple coarse-grained models to capture slow conformational transitions. Individually, each model describes one of the experimental structures; together, they approximate the complete energy surface. We demonstrate the method for the helix-to-sheet transition in Arc repressor N11L. We find that the transition involves the partial unfolding of the switch region, and rapid refolding into the alternate structure. Transient local unfolding is consistent with the low hydrogen exchange protection factors of the switch region. Also in agreement with experiment, the isomerization occurs independently of the global folding/dimerization transition.

Introduction

Structural studies have shown that transitions between different conformations of macromolecules and their complexes are important in biological processes. Examples range from allosteric transitions (Kern and Zuiderweg, 2003; Gunasekaran et al., 2004; e.g., in hemoglobin, see Perutz et al., 1998) to the structural transitions involved in the movement of motor proteins (Abrahams et al., 1994; Stock et al., 2000), or the conformational changes which occur upon binding of ligands to cell receptors (Bissantz, 2003). Experiments are now able to characterize the structure of populated states in great detail using X-ray diffraction and NMR spectroscopy. The rates of population exchange can also be probed by a variety of spectroscopic techniques, for example NMR spectroscopy (Akke, 2002; Lukin et al., 2003; Kay, 2005) or time-resolved optical spectroscopy (Henry et al., 1997). Nonetheless, it is also important to understand the transitions *between* the stable states. A description of the transitions should aid in understanding the mechanism and possibly allow us to interfere, for example by designing agonist ligands for receptors. Because the system is predominantly found in one of the stable states and not in the transition region, direct experimental investigations of the transitions between them are difficult, although the mechanism may be indirectly inferred from kinetic measurements as is done in protein folding Φ value analysis (Fersht et al., 1992). Whereas simulations can potentially fill in some of the details, most processes of interest occur on time (microsecond to second) and length (nanometer to micrometer) scales inaccessible to standard all-atom molecular dynamics simulations with transferable force fields (Karplus and McCammon, 2002); despite recent tour de force simulations of some very large systems, (Böckmann and Grübmuller, 2002; Aksimentiev and Schulten, 2005).

One approach to overcoming the problem of long time scales in simulation is to use coarse-grained models with simplified representations and energy functions (Tozzini, 2005), which incur a much lower computational cost than standard all-atom simulations. Due to the many degrees of freedom which have been integrated out, it is difficult, if not impossible, to parameterize transferable (i.e., sequence-based) energy functions for these models, at least using pair potentials (Vendruscolo and Domany, 1998; Moghaddam et al., 2005). Nonetheless, an energy function may be derived from the *structure* of each stable state, which is often known experimentally. A popular and successful class of such structure-based potentials comprises the elastic network models (Tirion, 1996), in which coarse particles representing the molecule at some level of detail are connected by harmonic springs; dynamical information can be inferred from normal mode analysis. Such descriptions have been applied to predicting isotropic thermal factors in protein crystal structures (Bahar et al., 1997), the movements of motor proteins (Zheng and Doniach, 2003; Navizet et al., 2004; Zheng and Brooks, 2005), ribosome motions (Tama et al., 2003; Wang et al., 2004), and the mechanical properties and assembly of virus capsids (Tama and Brooks, 2005; Rader et al., 2005). Elastic models have been particularly useful for the refinement and interpretation of low-resolution structural data by projecting structural changes onto low-frequency modes (Tama et al., 2003; Tama and Brooks, 2005; Ma, 2004). Going beyond this simple approach, a recent study has applied a perturbation analysis to the elastic network Hamiltonian to identify residues critical for transitions (Zheng et al., 2005). However, many molecular transitions will clearly not be described well by a harmonic model. This is especially true when the transition involves the polymeric character of the peptide chain, such as in protein folding, and when the energy landscape is locally rugged (Bryngelson and Wolynes, 1989). For these cases, a different class of *anharmonic* structure-based models has been used. In these so-called Gō models (Ueda et al., 1975), the protein is represented as a chain with attractive interactions between pairs of residues which interact in the native structure and repulsive interactions for all

*Correspondence: gerhard.hummer@nih.gov
[1]Lab address: http://www.niddk.nih.gov/intram/people/ghummer.htm

other residue pairs, producing an energy landscape with many local minima reflecting the formation and rupture of amino acid contacts and the rotation of backbone dihedral angles, but a funnel-like bias toward the native structure. These models have been very effective for the investigation of folding mechanisms (Clementi et al., 2000; Karanicolas and Brooks, 2002; Ollerenshaw et al., 2004; Hubner et al., 2004; Koga and Takada, 2001; Alm et al., 2002) and kinetics (Shimada and Shakhnovich, 2002; Karanicolas and Brooks, 2003a; Chavez et al., 2004; Henry and Eaton, 2004) and dimerization mechanisms (Yang et al., 2004; Levy et al., 2004, 2005a, 2005b), for example.

The remarkable success of the structure-based elastic network models and Gō models suggests that structural knowledge can be used to construct a reasonably good potential, at least in the vicinity of the experimental structure. However, both elastic network models and Gō-like models suffer from the limitation that the potential generally encodes only a single dominant minimum, corresponding to the structure used to derive it. Kim et al. (2002, 2005) have used a geometrical approach to generating intermediate structures. They define intermediate elastic potentials in which the distance matrix is a linear combination of those corresponding to initial and final structures. Another method of overcoming the single minimum problem, pioneered by Miyashita et al. (2003, 2005), is to recalculate the elastic network potential for intermediate structures along an assumed reaction path between the two dominant structures.

Here we present a scheme for building up an energy landscape encoding a number of dominant basins. We use structure-based potentials to model the energy surface in the vicinity of available experimental structures and present a method for merging the potentials that leads to a smooth transition between the different basins on the energy surface. In principle, this allows the assembly of energy landscapes encoding an arbitrary number of minima of known conformation. By stitching together structurally derived potentials that are accurate close to their respective minima, one aims to construct a globally reliable energy landscape. This method differs in several important respects from previous work combining different structure-based potentials. First, the individual structure-based energy functions are not limited to harmonic models, so that coarse polymer models (e.g., Gō models) may be used (as is done here) to model highly anharmonic transitions such as unfolding. Second, the system evolves on a single energy surface in which both stable conformers are minima, allowing ensembles of transition paths to be sampled by simulation. This may be useful if there is significant heterogeneity in the transition paths.

We have applied this approach to study a transition between helix and sheet structure in a protein context: the Arc repressor mutant N11L. It has been shown experimentally that the double mutation N11L,L12N causes the β sheet found in the structure of the wild-type Arc repressor dimer (Breg et al., 1990) to switch completely into a 3_{10} helix (Figure 1) (Cordes et al., 1999, 2003); the latter structure is commonly referred to as switch Arc. In this work, we refer to the wild-type structure as the β, or sheet form, and the switch mutant structure as the helical form, whereas the N-terminal region in which the structural change occurs will be denoted the switch region. Intriguingly, the intermediate single mutant N11L was found to populate both sheet and helix structures and to exchange between them on the NMR timescale under certain conditions (Cordes et al., 2000). The N11L mutant has therefore been described as a model for an evolutionary bridge between two structures. The transition between the helix and sheet forms of the protein would clearly not be captured by an elastic model, and a Gō model encoding one conformation will not find the other (shown below). However, by combining the Gō-like models derived from the experimental structures of each form, we are able to simulate transition paths and identify transition state structures between the two states.

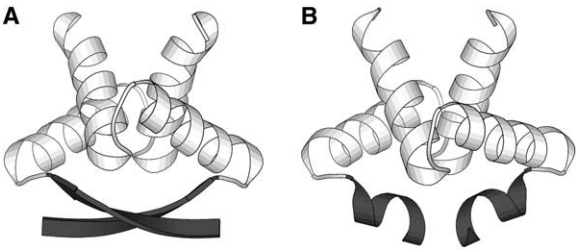

Figure 1. Experimental Arc Repressor Structures

Structures of the sheet (wild-type) and helix (switch) forms of Arc repressor are shown in (A) and (B), respectively. The region in which the structural transition occurs (residues 8–14) is shaded.

As the transition region is generally far from either experimental structure, it is arguably the most poorly approximated region of the energy landscape. To address this issue, we have also run simulations with additional non-Gō contacts. These do not qualitatively alter the results in this case, supporting the conclusions drawn from the double-Gō simulations.

Results and Discussion

How Can Multiple Energy Surfaces Be Combined?

Because individual structure-based models such as elastic network models and Gō models generally produce only a single dominant minimum, our approach is to combine two such models, one for each of the alternative experimental structures of Arc repressor. How should the two potential surfaces be combined in such a way as to preserve the shape of the energy surface in the vicinity of each minimum while giving a smooth transition between the two minima? Figure 2 illustrates some possible rules for combining two energy surfaces, for a pair of one-dimensional harmonic potentials (shown separately in Figure 2A). The simplest possibility, adding the two energies (Figure 2B), fails completely for harmonic potentials, as the result is a harmonic potential whose minimum is located midway between the minima of the original potentials! Although this effect may not be as severe for funnel-like potentials, the result for the harmonic case indicates that this approach could produce undesirable effects. Another simple alternative is to take the minimum of the two surfaces at every point in configuration space (Figure 2C). While this results in two minima at the correct locations, as desired, it also produces a cusp-like energy barrier which is unrealistic

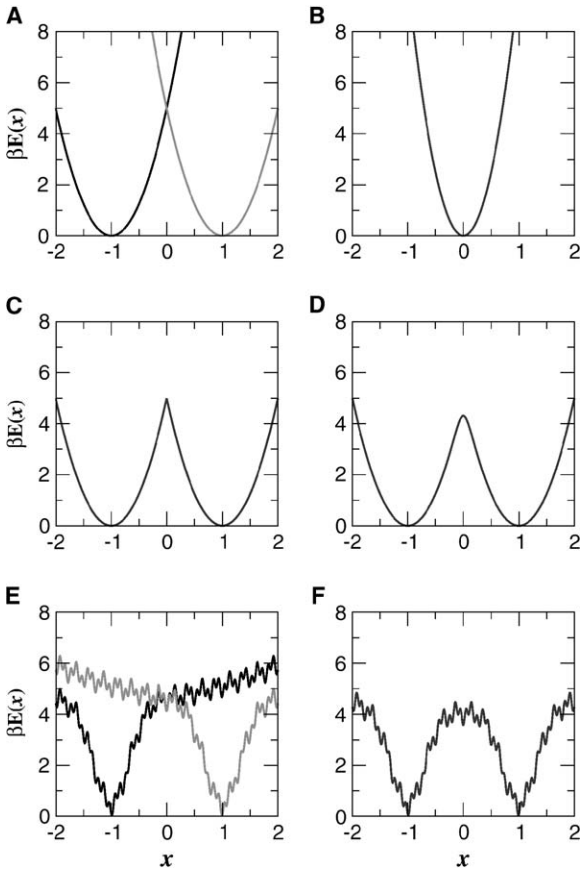

Figure 2. Possible Schemes for Combining Two Energy Functions

The two harmonic potentials $\beta E_{\pm}(x) = 5(x \pm 1)^2$, plotted separately as black and gray curves in (A), could be merged using either: (B) the sum $E_{sum}(x) = E_+(x) + E_-(x) - 10$ (the offset of −10 has been added for visualization only); (C) the minimum energy $E_{min}(x) = \min(E_+(x), E_-(x))$; or (D) the exponentially weighted sum (Equation 2), $E_{exp}(x) = -\beta^{-1}\ln(\exp(-\beta E_+(x)) + \exp(-\beta E_-(x)))$ (β = 1 in the figure). The exponentially weighted sum in (D) is equally applicable to anharmonic potentials, such as the two cartoon energy landscapes shown separately in (E) and merged in (F). This would be the situation for Gō models, for instance.

and unsuitable for simulations requiring differentiable energies, such as molecular dynamics or Langevin dynamics. Our proposed method is based on an exponential Boltzmann weighting of the two surfaces, given by Equation 2 (see Experimental Procedures) (Hummer et al., 1997): the resulting energy surface (Figure 2D) is essentially identical to the separate models in the vicinity of the minima in the original energy functions, but has a smooth, continuous barrier in the transition region. In the language of statistical mechanics, we add up the partition functions corresponding to the individual energy surfaces rather than the energies themselves; that is to say, we pool the accessible states defined by the individual energy functions. In principle, as many surfaces as desired could be combined using this method (Equation 2). For harmonic potentials, the equilibrium properties of such mixed models may be calculated analytically. Although the method was illustrated for harmonic potentials in Figure 2D, it is equally applicable to anharmonic potentials such as Gō models. For example, in Figure 2E are shown two cartoon one-dimensional funnel landscapes with some local ruggedness, which can be mixed with our exponential weighting scheme to give the combined potential plotted in Figure 2F.

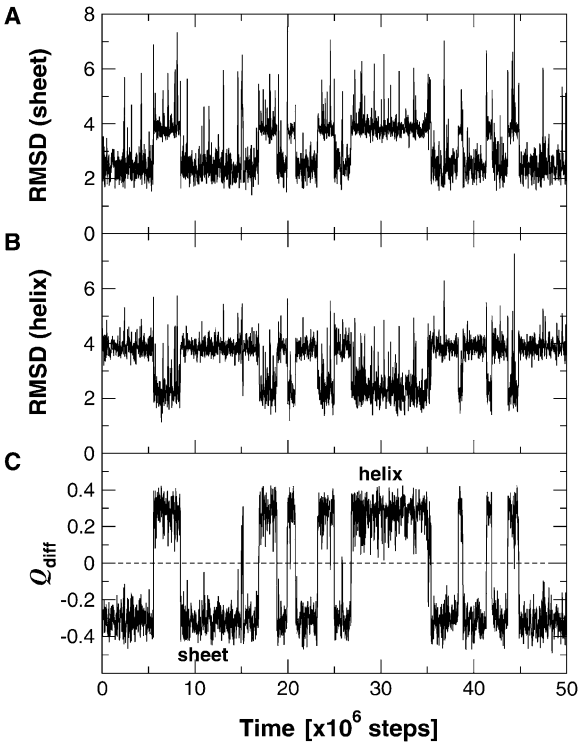

Figure 3. Equilibrium Helix-Sheet Transitions of Mixed Gō Models of Arc

(A) and (B) show the global C_α root mean square deviations (rmsd) from the experimental structures of the (A) sheet and (B) helix forms of Arc (in Å) for simulations of Arc repressor in the mixed Gō potential. In (C), the trajectory is projected onto the coordinate Q_{diff}, defined as the difference between the fraction of native contacts involving the switch region in the helix form of Arc (Q_h) and the fraction of native contacts involving the switch region in the sheet form of Arc (Q_s).

Double-Gō Model for Arc Repressor

Gō-like models were derived from the NMR structures of wild-type Arc (Protein Data Bank code 1ARR; Figure 1A) and switch Arc (PDB code 1QTG; Figure 1B) using a standard prescription (Karanicolas and Brooks, 2002). A number of small modifications were made (described in Experimental Procedures), in particular to make the backbone potential generic, that is, independent of the folded structures. The most significant change was the introduction of a fully transferable pseudoangle potential for the C_α-C_α-C_α angles.

The separate Gō-like models of sheet and helix Arc each remain folded over the range of temperatures simulated (280–320 K) with a C_α root mean square deviation (rmsd) from the experimental structures of around 2 Å. When the two Gō potentials are combined using Equation 2, the protein exchanges frequently between the sheet and helix forms at equilibrium. The rmsd of a typical equilibrium trajectory at 300 K from each stable structure is shown in Figures 3A and 3B. The protein converts between the two alternate conformers in an essentially two-state fashion. Note that there are frequent

fluctuations in rmsd from each state that do not result in transitions. Similar fluctuations, attributable to the switch region, also occur in the separate Gō models for the sheet and helix forms. These rapid fluctuations suggest that the switch region is a less stable element of structure compared with the rest of the protein. Protection factors from hydrogen exchange experiments on both sheet and helix forms of the proteins show that the switch region is less stable than the main body of the protein, consistent with the fluctuations seen in our model (Cordes et al., 2003). The flexibility of this region of the protein may be related to its function in binding DNA; certain residues, for example, Phe 10, are known to flip out in order to bind DNA, and NMR dynamics suggests that a preequilibrium exists between the two conformers (Nooren et al., 1999).

The global folding transition of the mixed Gō model occurs only at a much higher temperature ($\approx$360 K), and is therefore independent of the local helix-to-sheet isomerization; this is consistent with the experimental interpretation that the structural change in Arc N11L occurs independently of the overall folding of the protein (Cordes et al., 1999). We find that global folding/unfolding transitions at 360 K are tightly coupled with the binding of the two monomers (see Supplemental Data available with this article online), in agreement with experiment (Milla and Sauer, 1994; Milla et al., 1995) and with earlier simulation studies using a different Gō-like model (Levy et al., 2004, 2005a).

In Figure 3C, the same trajectory segment has been projected onto a reaction coordinate Q_{diff}, which separates the two stable states very well. This coordinate is defined as the difference between the fraction of native contacts involving the switch region in helical Arc (Q_h) and the fraction of native contacts involving the switch region in the sheet form of Arc (Q_s). We also find that this reaction coordinate is suitable for identifying transition states, by applying a Bayesian approach to identifying reactive states (Hummer, 2004; Best and Hummer, 2005); we identify transition states as those configurations with the highest probability that trajectories passing through them are reactive, that is, connect the sheet conformations ($Q_{diff} = -0.3$) to the helical conformations ($Q_{diff} = 0.3$). This can be quantified by the conditional probability of being on a transition path given a value of Q_{diff}, $p(TP|Q_{diff})$. The largest value of $p(TP|Q_{diff})$, corresponding to the most likely transition states, is $\approx$0.4 for $Q_{diff} = -0.034$ (Figure S1). Because this is close to the maximum possible value of 0.5 for diffusive processes (Hummer, 2004), transition states may be derived using the Q_{diff} reaction coordinate with confidence. Using this value of Q_{diff} as a dividing surface, we estimate the ratio of the populations of the protein in sheet and helix conformers to be approximately 67:33. This balance is obtained from mixing the two potentials using Equation 2 with the offsets ϵ_i set to zero, but ϵ_i could in principle be adjusted to account for differences in stability, or to tune the stability to favor one of the states as is often done experimentally. From the mean residence times in each state, we obtain the rate for conversion of sheet to helix $k_{s\rightarrow h}$ as 20.6 (2.3) μs^{-1} and the reverse rate $k_{h\rightarrow s}$ as 41.7 (4.7) μs^{-1}. These rates were calculated from simulations with a damping coefficient γ approximately three orders of magnitude smaller than that commonly used for liquid water, in order to speed up the rate of transitions. Although there is no guarantee for such simplified models that the calculated rates should correspond to experiment, they are in fact about three orders of magnitude faster than those measured from analysis of NMR line shapes (Cordes et al., 1999).

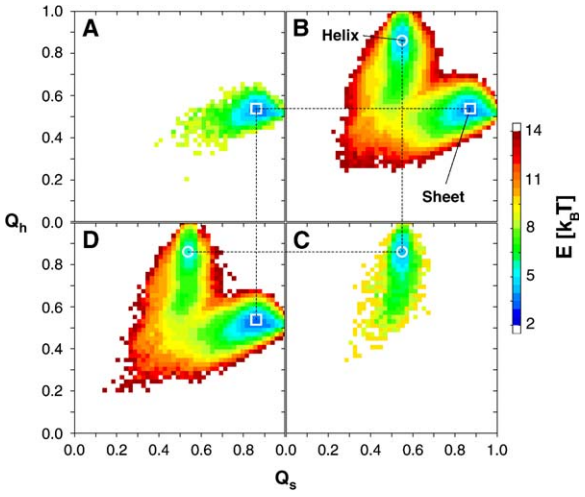

Figure 4. Free Energy Surfaces for Arc Repressor Models

Two-dimensional potentials of mean force are shown as a function of the fraction of sheet native contacts Q_s and helix native contacts Q_h involving the switch region, obtained from equilibrium simulations of Arc at 300 K using (A) the sheet potential, (B) the mixed potential, and (C) the helix potential. Note that to facilitate comparison, the free energies in (A) and (C) have been shifted using the relative populations of the sheet and helix structures in the ensemble in (B). The corresponding free energy surface for a mixed potential incorporating additional non-Gō contacts is shown in (D). The position of the sheet minimum in (A), (B), and (D) is indicated by a white square, and the position of the helix minimum in (B), (C), and (D) by a white circle, for purposes of comparison; broken lines link the minima in the different plots.

Some insight into the shape of the energy surface resulting from the mixing of the two models, and its relation to those arising from the separate models, may be obtained by comparing free energy surfaces as a function of suitable reaction coordinates: we choose Q_s and Q_h, because contact-based reaction coordinates are often good coordinates for Gō-like models when the system is relatively unfrustrated (Clementi et al., 2000; Best and Hummer, 2005); they may break down for more rugged potentials or for larger proteins. The potential of mean force in terms of these coordinates is shown in Figure 4B. The free energy surface has two principal minima: one corresponding to the sheet structure ($Q_s = 0.86$, $Q_h = 0.54$) and the other to the helix structure ($Q_s = 0.53$, $Q_h = 0.87$), with a broad transition region connecting them. In this two-dimensional projection, the saddle in the free energy surface consists of structures in which the switch region is essentially unfolded with respect to either the helix or sheet structures (although the remainder of the protein remains well folded). Analogous free energy surfaces for the separate sheet and helix potentials are shown in Figures 4A and 4C; in each of these, as expected, there is only a single minimum corresponding to the structure used to derive the potential, and the alternate conformer is not populated. As expected from the mixing rule, the energy surface

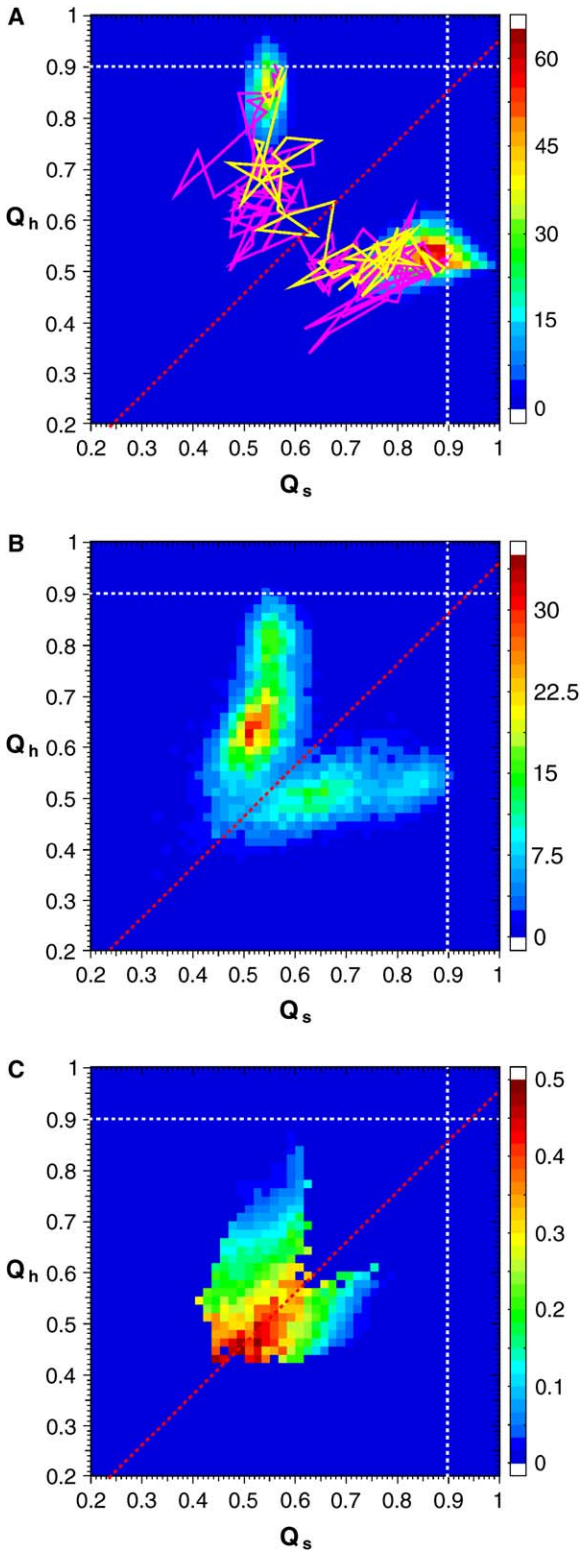

Figure 5. Two-Dimensional Bayesian Analysis of Transition Paths between the Sheet and Helix Forms of Arc

The equilibrium probability density $p_{eq}(Q_s,Q_h)$ is plotted in (A), showing the slight preference for the sheet conformer. Transition paths are defined as those segments of equilibrium trajectories which connect the two regions defined by $Q_s > 0.9$ and $Q_h > 0.9$ (the borders of these regions are indicated by broken white lines in each panel); two examples of such transition paths are superimposed on the equilibrium probability density in (A) (yellow and magenta lines). (B) shows the conditional probability of the reaction coordinates for transition paths only, $p(Q_s,Q_h|TP)$. In (C), the most reactive states are identified using $p(TP|Q_s,Q_h)$. In all panels, the locus of the transition state identified from the one-dimensional reaction coordinate $Q_{diff} = Q_h - Q_s = -0.034$ is shown with a broken red line.

arising from the combined potential in Figure 4B is essentially identical to the free energy surfaces for the separate potentials in Figures 4A and 4C in regions close to either minimum. In addition, combining the potentials results in a smooth saddle connecting the two minima.

Mechanism of Helix/Sheet Interconversion

Figure 4B is suggestive of a mechanism for interconversion of the sheet and helix forms of Arc: the minimum free energy path would involve initially unfolding one structure and then forming the new structure, following an L-shaped path on the projection. However, the broad saddle lying along the diagonal suggests that there may also be pathways in which formation of the new structure occurs concomitantly with the loss of the old. We have therefore studied the transition paths between the two states in more detail using a Bayesian approach (Hummer, 2004; Best and Hummer, 2005) (Figure 5). For this two-dimensional analysis, we identify transition paths as being segments of equilibrium trajectories connecting the two regions of Q space defined by $Q_h > 0.9$ and $Q_s > 0.9$ without recrossing (Hummer, 2004). The equilibrium probability density $p_{eq}(Q_s,Q_h)$ in Figure 5A has the expected two-state appearance, with a slightly higher density for the favored sheet form. The conditional probability density $p(Q_s,Q_h|TP)$ obtained from a histogram of the transition paths shows that the most probable route for transitions is indeed via the approximately L-shaped path inferred from the free energy surface. Moreover, Figures 4B and 5A indicate that there are no significantly populated intermediate structures which would complicate the analysis; this is also consistent with the two-state interpretation of the experimental data (Cordes et al., 1999). Finally, we are able to use the Bayesian relation,

$$p(TP|Q_s,Q_h) = p(Q_s,Q_h|TP)p(TP)/p_{eq}(Q_s,Q_h), \qquad (1)$$

with $p(TP)$ being the equilibrium probability of being on a transition path, to estimate the probability of being on a transition path given certain values of the reaction coordinates, $p(TP|Q_s,Q_h)$. This indicates that the most reactive states (transition states), being those with the highest values of $p(TP|Q_s,Q_h)$, lie approximately along the diagonal. Furthermore, it explains why the coordinate $Q_{diff} = Q_h - Q_s$ is a good one-dimensional reaction coordinate: it is orthogonal to the dividing surface (stochastic separatrix) (Berezhkovskii and Szabo, 2005). In this case, it appears that the two-dimensional coordinates Q_s, Q_h do not represent a significant improvement over the one-dimensional coordinate Q_{diff} in terms of identifying transition states.

The one-dimensional free energy profile along the coordinate Q_{diff} is shown in Figure 6A, revealing minima for the helix and sheet structures and an intervening barrier. Because Q_{diff} is a good reaction coordinate, it may also be used to characterize the structures along the

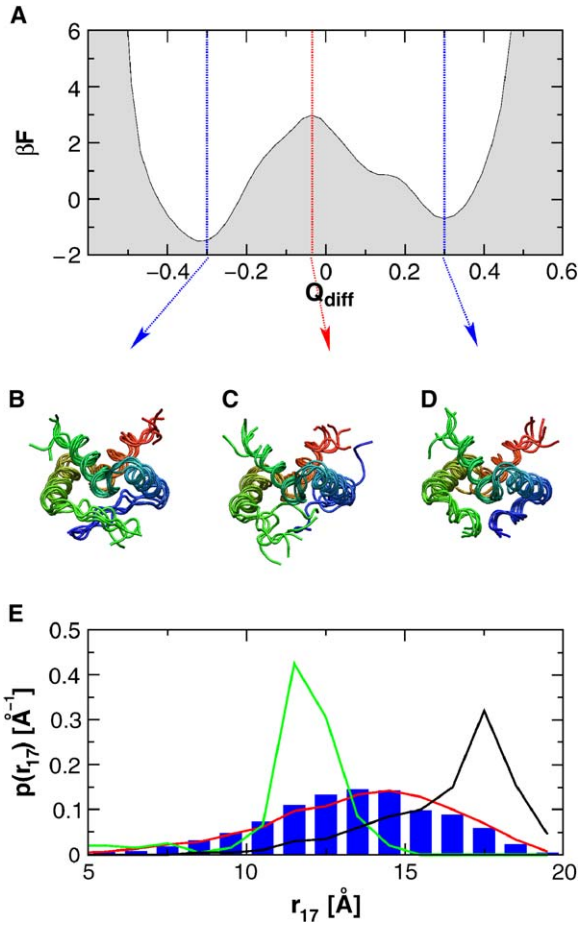

Figure 6. Identifying Transition States for the Helix-Sheet Transition of Arc

(A) Potential of mean force along the reaction coordinate $Q_{diff} = Q_h - Q_s$.

(B–D) Structures corresponding to selected values of Q_{diff} have been used to illustrate the transition: (B) $Q_{diff} \approx -0.3$ (sheet or wild-type structure); (C) $Q_{diff} \approx -0.034$ (transition state structures); and (D) $Q_{diff} \approx 0.3$ (helical or switch structure). In each case, the five structures closest to the chosen value of the reaction coordinate are shown.

(E) The distribution of end-to-end lengths of the switch region (residues 8–14) in the transition state (blue histogram) is compared with that in the helical form (green curve), the sheet form (black curve), and the distribution of end-to-end lengths of a 7 residue peptide with no attractive interactions (red curve).

transition pathway. By selecting structures from the trajectories corresponding to certain values of the reaction coordinate, we can identify structures representative of the sheet, helix, and transition state ensembles (Figures 6B, 6D, and 6C, respectively). Visually, the switch region is disordered in the transition state, consistent with the interpretation in terms of local unfolding above. A simple means of quantifying the disorder is via the distribution of end-to-end distances for the switch region. Figure 6E shows the distribution of distances between residues 8 and 14, which bound the switch region. The blue histogram of distances from transition state configurations is very broad, and clearly differs from the narrow distance distributions for the helix (green) and sheet (black) states. A random chain version of our model was generated by running simulations of a seven residue peptide with the same potential as the Gō model for residues 8–14, but with all attractive contacts turned off. The distribution of end-to-end distances for the random chain overlays quite well with the transition state distribution of lengths, being only slightly shifted to longer lengths (by about 0.3 Å). While the end-to-end distance is clearly a simplified description, this correspondence suggests that a random chain is a reasonable approximation for the transition state. We can use this finding to suggest some perturbations that should alter the transition rate. For example, adding a low concentration of denaturant may stabilize the transition state and thus increase the rate of the reaction; however, higher denaturant concentrations (still below those required to unfold the protein) may result in the unfolded transition state becoming a stable intermediate, slowing the transition rate again. Similarly, incorporation of glycine into the switch region would destabilize both sheet and helix relative to the transition state, and may also increase the rate.

Although the transition appears essentially two-state, there is a slight shoulder on the helical side of the free energy barrier in Figure 6A around $Q_{diff} \approx 0.15$, suggesting an unstable intermediate; although this is hard to see in the two-dimensional equilibrium probability density (Figure 5A) due to its low overall population, it shows up quite clearly in the two-dimensional probability density for transition paths (Figure 5B) as a peak at $Q_h \approx 0.65$, $Q_s \approx 0.52$. A frequent feature of structures matching $Q_{diff} \approx 0.15$ is the presence of only a single helix. This suggests that the two helices can unfold sequentially during helix-to-sheet conformational transitions. In contrast, no structures with only one strand of the sheet present are observed, presumably because almost all of the sheet contacts are between the two monomers, in contrast to the situation for helices. What is clear from both one- and two-dimensional analysis, however, is that both helices unfold before the transition state is reached.

We note here that our finding of two-state transitions for this system was not predetermined by the model: for example, single Gō models for folding have been found to fold via intermediates (Clementi et al., 2000; Karanicolas and Brooks, 2003b), without the intermediate being explicitly incorporated into the energy function. However, if the structure of a stable intermediate were known from experiment, this could be incorporated into the energy function as an additional structure-based potential. A better description of nonnative intermediates could also be obtained by using a more physics-based potential function than a Gō model (for example, by including some nonnative contact energy as described below).

Influence of Nonnative Contacts

Gō models are expected to be a reasonable approximation close to the structure used to derive them, and many folding studies have suggested that they are a sufficiently good model away from this structure to predict folding rates (Chavez et al., 2004; Henry and Eaton, 2004) and, more qualitatively, mechanisms (Shoemaker et al., 1999; Muñoz and Eaton, 1999; Clementi et al., 2000; Koga and Takada, 2001; Alm et al., 2002; Henry and Eaton, 2004). Nonetheless, the regions of the mixed

energy landscape in the vicinity of the transition region should be more poorly approximated, given that they are far from either stable structure. How important are the missing non-Gō interactions (those contacts not formed in the two structures used to construct the double-Gō potential) for defining the transition region between the two states?

To answer this question for Arc repressor, we have built an identical double-Gō potential, but the repulsive terms for non-Gō contacts have been substituted with an attractive potential, with a well depth of 40% of the equivalent depth for native contacts. In general, we find that the addition of non-Gō contacts has only a small influence on the results. The sheet form of the protein is slightly stabilized relative to the pure Gō models (the ratio of sheet:helix shifts to 86:14), and the single-helix intermediate is slightly stabilized. However, the overall shape of the two-dimensional free energy surface is essentially unchanged (Figure 4D), and the transition state retains its unfolded character. Whereas non-Gō contacts may be important for other systems, we note that the method of mixing potentials presented here ensures that the potential for the transition region will be at least as good as that for any of the underlying single-minimum models.

Conclusion

Thanks to advances in structural biology, there are now many techniques for determining structures of stable conformations of proteins and other macromolecules under suitable conditions. By themselves, however, these structures do not reveal the mechanism by which such states interconvert; perhaps only single-molecule experiments could be used to obtain *direct* information on transition paths. This is a natural niche for simulations with coarse models: experimental structures can be used to derive coarse potentials, while the method presented in this work can be used to splice together a number of potentials derived from different experimental structures to create a more faithful representation of the energy surface. Mechanisms obtained directly from such simulations may also be tested against experimental information, such as Φ values (Fersht et al., 1992).

Consistent with experimental evidence, we find that the transition between the sheet and helix forms of Arc repressor N11L is two-state, and does not require the unfolding or dissociation of the Arc dimer. Instead, the transition occurs by local unfolding of the switch region followed by rapid refolding.

Experimental Procedures

Mixed Potential

Consider two energy functions (e.g., for elastic network models or Gō models) $E_1(\mathbf{R})$ and $E_2(\mathbf{R})$ with global minima at two conformers $\mathbf{R}_1$ and $\mathbf{R}_2$, respectively. The corresponding partition functions are $Z_1 = \int d\mathbf{R} \exp(-\beta E_1(\mathbf{R}))$ and $Z_2 = \int d\mathbf{R} \exp(-\beta E_2(\mathbf{R}))$. To combine the models such that their respective minima are retained, we add their partition functions $Z = Z_1 + Z_2$ (Hummer et al., 1997). Expressed in terms of a new potential function, and for N potential surfaces $E_i(\mathbf{R})$, one obtains:

$$\exp(-\beta E(\mathbf{R})) = \sum_{i=1}^{N} \exp(-\beta(E_i(\mathbf{R}) + \varepsilon_i)). \quad (2)$$

In this expression, we choose a mixing parameter $\beta = 1/k_B T$ and ϵ_i are offsets to balance the relative free energies of the different minima. Energy surfaces $E_i(\mathbf{R})$ can be taken from elastic models, Gō models, or all-atom transferable potentials.

This method was implemented in the molecular dynamics code CHARMM (Brooks et al., 1983). A separate copy of the executable is run for each potential, and the energies and forces are averaged according to Equation 2 at each time step using the portable MPI interface (Snir et al., 1998); the technical details of the MPI implementation are as previously described (Best and Vendruscolo, 2004).

Generation of Gō-like Potentials

Models of the sheet and helix forms of Arc repressor were based on the NMR structures PDB codes 1ARR (Breg et al., 1990) and 1QTG (Cordes et al., 1999) of the wild-type and switch proteins, respectively. Only those residues (8–48) which were found by NMR to be structured were used in building the models, though the original residue numbering is retained here for consistency. A standard procedure (Karanicolas and Brooks, 2002) was used to build a Gō-like model of each protein in which each amino acid is represented by a single bead at the C_α carbon position. The potential consists of a transferable dihedral potential, a harmonic potential for angles with a minimum at the native angle, an attractive term (Miyazawa and Jernigan, 1996) for contacts which are formed in the experimental structure (native contacts), and a repulsive potential for the remaining (nonnative) contacts. Bonds are constrained to their lengths in the native structure using SHAKE (Ryckaert et al., 1977). Although the respective experimental structures were used to derive the contact lists, the sequence-dependent part of the model was obtained from the same sequence, that of the N11L mutant.

In order to make the model compatible with the mixed potential described above, the following changes were made. The length of each C_α-C_α pseudobond was taken as the average over the two structures; this is reasonable, as the bond lengths are narrowly distributed about 3.81 Å (in fact setting all bond lengths to 3.81 Å hardly affects the results). Also, the C_α-C_α-C_α pseudoangles differ significantly for α and β structure (being approximately 90° and 120°–130°, respectively). Because the angles are described by a relatively stiff harmonic potential in the standard model (Karanicolas and Brooks, 2002), the overlap of the angle distributions in the helix and sheet forms of the protein would be very poor. The pseudoangles were therefore each described by a generic double well potential, derived from a statistical analysis of the TOP500 set of protein structures (Lovell et al., 2003), which was taken to be the same for all angles. We note that this analysis also supports the use of a common angle potential for all pseudoangles (i.e., independent of the central residue).

$$\exp(-\gamma E(\theta)) = \exp(-\gamma(k_\alpha(\theta - \theta_\alpha)^2 + \varepsilon_\alpha)) + \exp(-\gamma k_\beta(\theta - \theta_\beta)^2) \quad (3)$$

The values of the constants are $\gamma = 0.1$ mol.kcal^{-1}, $\epsilon_\alpha = 4.3$ kcal.mol^{-1}, $\theta_\alpha = 1.60$ rad, $\theta_\beta = 2.27$ rad, $k_\alpha = 106.4$ kcal.mol^{-1}.rad^{-2}, and $k_\beta = 26.3$ kcal.mol^{-1}.rad^{-2}. With a potential defined in this way, each angle can independently isomerize between α-like and β-like values.

The increased entropy of the unfolded state resulting from the more flexible angles significantly destabilizes the proteins; to compensate, all attractive contact energies were scaled by a factor of 1.7. In addition, the contact lists were consolidated as follows: the attractive contacts within the nonswitch part of the protein (residues 15–48) were taken from the sheet model, whereas contacts within the switch region (residues 8–14), and between this region and the nonswitch region, were taken from the respective (helix or sheet) models. This choice is motivated by the fact that the structure of the nonswitch region is essentially identical for the sheet and helix structures, differing by less than 1.0 Å for each of residues 15–46 when the nonswitch regions are superimposed by least-squares alignment. The repulsive radius of each atom was chosen to be the minimum from the two models, to minimize hard-core overlap. The potentials for the sheet and helix forms were combined using Equation 2, setting the offsets ϵ_i to zero for each Gō model.

Models with non-Gō contacts were constructed as follows: the original model was modified by treating interactions not present in the template structure as attractive, instead of repulsive. The interactions were treated with a 12–10 potential with a minimum radius of $\sigma = 5.5$ Å, and a well depth weighted by the same Miyazawa-Jernigan contact energies (Miyazawa and Jernigan, 1996) as the

native contacts in the Karanicolas model (Karanicolas and Brooks, 2002), but scaled to 40% of the native value.

Simulations

Langevin dynamics simulations were run in CHARMM using the Brooks, Brünger, and Karplus algorithm (Brooks et al., 1984; Pastor et al., 1988) with a friction coefficient of 0.1 ps^{-1} and a time step of 15 fs. This friction coefficient is far smaller than values mimicking the friction of water. Typical values approximating water friction are in the range of 50–100 ps^{-1} (Pastor and Karplus, 1989; Snow et al., 2002). The low value of the friction has been chosen in order to increase the rate of transitions in the equilibrium simulations performed here, based on transition rates in a similar Gō model for folding (Best and Hummer, 2005). All bonds were fixed at their native lengths using SHAKE (Ryckaert et al., 1977). The aggregate simulation time was 4×10^8 steps (6.0 μs), including 156 transitions.

Supplemental Data

Supplemental data, including a figure, can be found with this article online at http://www.structure.org/cgi/content/full/13/12/1755/DC1/.

Acknowledgments

We would like to thank W.A. Eaton and A. Szabo for helpful comments on the manuscript. This research was supported by the Intramural Research Program of the NIH, NIDDK.

Received: June 29, 2005
Revised: August 2, 2005
Accepted: August 10, 2005
Published: December 13, 2005

References

Abrahams, J.P., Leslie, A.G.W., Lutter, R., and Walker, J.E. (1994). Structure at 2.8-Angstrom resolution of F_1-ATPase from bovine heart mitochondria. Nature *370*, 621–628.

Akke, M. (2002). NMR methods for characterizing microsecond to millisecond dynamics in recognition and catalysis. Curr. Opin. Struct. Biol. *12*, 642–647.

Aksimentiev, A., and Schulten, K. (2005). Imaging α-hemolysin with molecular dynamics: ionic conductance, osmotic permeability, and the electrostatic potential map. Biophys. J. *88*, 3745–3761.

Alm, E., Morozov, A.V., Kortemme, T., and Baker, D. (2002). Simple physical models connect theory and experiment in protein folding kinetics. J. Mol. Biol. *322*, 463–476.

Bahar, I., Atilgan, A.R., and Erman, B. (1997). Direct evaluation of thermal fluctuations in proteins using a single-parameter harmonic potential. Fold. Des. *2*, 173–181.

Berezhkovskii, A., and Szabo, A. (2005). One-dimensional reaction coordinates for diffusive activated rate processes in many dimensions. J. Chem. Phys. *122*, 14503.

Best, R.B., and Hummer, G. (2005). Reaction coordinates and rates from transition paths. Proc. Natl. Acad. Sci. USA *102*, 6732–6737.

Best, R.B., and Vendruscolo, M. (2004). Determination of ensembles of protein structures consistent with NMR order parameters. J. Am. Chem. Soc. *126*, 8090–8091.

Bissantz, C. (2003). Conformational changes of G protein-coupled receptors during their activation by agonist binding. J. Recept. Signal Transduct. Res. *23*, 123–153.

Böckmann, R.A., and Grübmuller, H. (2002). Nanoseconds molecular dynamics simulation of primary mechanical energy transfer steps in F_1-ATP synthase. Nat. Struct. Biol. *9*, 198–202.

Breg, J.N., van Opheusden, J.H.J., Burgering, M.J.M., Boelens, R., and Kaptein, R. (1990). Structure of Arc repressor in solution: evidence for a family of β-sheet DNA-binding proteins. Nature *346*, 586–589.

Brooks, B.R., Bruccoleri, R.E., Olafson, B.D., States, D.J., Swaminathan, S., and Karplus, M. (1983). CHARMM: a program for macromolecular energy, minimization, and dynamics calculations. J. Comput. Chem. *4*, 187–217.

Brooks, C.L., III, Brünger, A., and Karplus, M. (1984). Stochastic boundary-conditions for molecular dynamics simulations of ST2 water. Chem. Phys. Lett. *105*, 495–500.

Bryngelson, J.D., and Wolynes, P.G. (1989). Intermediates and barrier crossing in a random energy model (with applications to protein folding). J. Phys. Chem. *93*, 6902–6915.

Chavez, L.L., Onuchic, J.N., and Clementi, C. (2004). Quantifying the roughness on the free energy landscape: entropic bottlenecks and protein folding rates. J. Am. Chem. Soc. *126*, 8426–8432.

Clementi, C., Nymeyer, H., and Onuchic, J.N. (2000). Topological and energetic factors: what determines the structural details of the transition state ensemble and "en-route" intermediates for protein folding? An investigation for small globular proteins. J. Mol. Biol. *298*, 937–953.

Cordes, M.H.J., Walsh, N.P., McKnight, C.J., and Sauer, R.T. (1999). Evolution of a protein fold in vitro. Science *284*, 325–327.

Cordes, M.H.J., Burton, R.E., Walsh, N.P., McKnight, C.J., and Sauer, R.T. (2000). An evolutionary bridge to a new protein fold. Nat. Struct. Biol. *7*, 1129–1132.

Cordes, M.H.J., Walsh, N.P., McKnight, C.J., and Sauer, R.T. (2003). Solution structure of switch Arc, a mutant with 3_{10} helices replacing a wild-type β-ribbon. J. Mol. Biol. *326*, 899–909.

Fersht, A.R., Matouschek, A., and Serrano, L. (1992). The folding of an enzyme. 1. Theory of protein engineering analysis of stability and pathway of protein folding. J. Mol. Biol. *224*, 771–782.

Gunasekaran, K., Ma, B.Y., and Nussinov, R. (2004). Is allostery an intrinsic property of all dynamic proteins? Proteins *57*, 433–443.

Henry, E.R., and Eaton, W.A. (2004). Combinatorial modeling of protein folding kinetics: free energy profiles and rates. Chem. Phys. *307*, 163–185.

Henry, E.R., Jones, C.M., Hofrichter, J., and Eaton, W.A. (1997). Can a two-state MWC allosteric model explain hemoglobin kinetics? Biochemistry *36*, 6511–6528.

Hubner, I.A., Oliveberg, M., and Shakhnovich, E.I. (2004). Simulation, experiment, and evolution: understanding nucleation in protein S6 folding. Proc. Natl. Acad. Sci. USA *101*, 8354–8359.

Hummer, G. (2004). From transition paths to transition states and rate coefficients. J. Chem. Phys. *120*, 516–523.

Hummer, G., Pratt, L.R., and García, A.E. (1997). Multistate Gaussian model for electrostatic solvation free energies. J. Am. Chem. Soc. *119*, 8523–8527.

Karanicolas, J., and Brooks, C.L., III. (2002). The origins of asymmetry in the folding transition states of protein L and protein G. Protein Sci. *11*, 2351–2361.

Karanicolas, J., and Brooks, C.L., III. (2003a). The structural basis for biphasic kinetics in the folding of the WW domain from a formin-binding protein: lessons for protein design? Proc. Natl. Acad. Sci. USA *100*, 3954–3959.

Karanicolas, J., and Brooks, C.L., III. (2003b). Improved Gō-like models demonstrate the robustness of protein folding mechanisms towards non-native interactions. J. Mol. Biol. *334*, 309–325.

Karplus, M., and McCammon, J.A. (2002). Molecular dynamics simulations of biomolecules. Nat. Struct. Biol. *9*, 646–652.

Kay, L.E. (2005). NMR studies of protein structure and dynamics. J. Magn. Reson. *173*, 193–207.

Kern, D., and Zuiderweg, E.R.P. (2003). The role of dynamics in allosteric regulation. Curr. Opin. Struct. Biol. *13*, 748–757.

Kim, M.K., Chirikjian, G.S., and Jernigan, R.L. (2002). Elastic models of conformational transitions in macromolecules. J. Mol. Graph. Model. *21*, 151–160.

Kim, M.K., Jernigan, R.L., and Chirikjian, G.S. (2005). Rigid-cluster models of conformational transitions in macromolecular machines and assemblies. Biophys. J. *89*, 43–55.

Koga, N., and Takada, S. (2001). Roles of native topology and chain-length scaling in protein folding: a simulation study with a Gō-like model. J. Mol. Biol. *313*, 171–180.

Levy, Y., Wolynes, P.G., and Onuchic, J.N. (2004). Protein topology determines binding mechanism. Proc. Natl. Acad. Sci. USA *101*, 511–516.

Levy, Y., Cho, S.S., Onuchic, J.N., and Wolynes, P.G. (2005a). A survey of flexible protein binding mechanisms and their transition states using native topology based energy landscapes. J. Mol. Biol. *346*, 1121–1145.

Levy, Y., Cho, S.S., Onuchic, J.N., and Wolynes, P.G. (2005b). Symmetry and frustration in protein energy landscapes: a near degeneracy resolves the Rop dimer-folding mystery. Proc. Natl. Acad. Sci. USA *102*, 2373–2378.

Lovell, S.C., Davis, I.W., Arendall, W.B., III, de Bakker, P.I., Word, J.M., Prisant, M.G., Richardson, J.S., and Richardson, D.C. (2003). Structure validation by Cα geometry: ϕ, ψ and Cβ deviation. Proteins *50*, 437–450.

Lukin, J.A., Kontaxis, G., Simplaceanu, V., Yuan, Y., Bax, A., and Ho, C. (2003). Quaternary structure of hemoglobin in solution. Proc. Natl. Acad. Sci. USA *100*, 517–520.

Ma, J.P. (2004). New advances in normal mode analysis of supermolecular complexes and applications to structural refinement. Curr. Protein Pept. Sci. *5*, 119–123.

Milla, M.E., and Sauer, R.T. (1994). P22 Arc repressor—folding kinetics of a single-domain dimeric protein. Biochemistry *33*, 1125–1133.

Milla, M.E., Brown, B.M., Waldburger, C.D., and Sauer, R.T. (1995). P22 Arc repressor: transition state properties inferred from mutational effects on the rates of protein unfolding and refolding. Biochemistry *34*, 13914–13919.

Miyashita, O., Onuchic, J.N., and Wolynes, P.G. (2003). Nonlinear elasticity, proteinquakes, and the energy landscapes of functional transitions in proteins. Proc. Natl. Acad. Sci. USA *100*, 12570–12575.

Miyashita, O., Wolynes, P.G., and Onuchic, J.N. (2005). Simple energy landscape model for the kinetics of functional transitions in proteins. J. Phys. Chem. B *109*, 1959–1969.

Miyazawa, S., and Jernigan, R.L. (1996). Residue-residue potentials with a favourable contact pair term and an unfavourable high packing density term, for simulation and threading. J. Mol. Biol. *256*, 623–644.

Moghaddam, M.S., Shimizu, S., and Chan, H.S. (2005). Temperature dependence of three-body hydrophobic interactions: potential of mean force, enthalpy, entropy, heat capacity and nonadditivity. J. Am. Chem. Soc. *127*, 303–316.

Muñoz, V., and Eaton, W.A. (1999). A simple model for calculating the kinetics of protein folding from three-dimensional structures. Proc. Natl. Acad. Sci. USA *96*, 11311–11316.

Navizet, I., Lavery, R., and Jernigan, R.L. (2004). Myosin flexibility: structural domains and collective vibrations. Proteins *54*, 384–393.

Nooren, I.M.A., Rietveld, A.W.M., Melacini, G., Sauer, R.T., Kaptein, R., and Boelens, R. (1999). The solution structure and dynamics of an Arc repressor mutant reveal premelting conformational changes related to DNA binding. Biochemistry *38*, 6035–6042.

Ollerenshaw, J.E., Kaya, H., Chan, H.S., and Kay, L. (2004). Sparsely populated folding intermediates of the Fyn SH3 domain: matching native-centric essential dynamics and experiment. Proc. Natl. Acad. Sci. USA *101*, 14748–14753.

Pastor, R.W., and Karplus, M. (1989). Inertial effects in butane stochastic dynamics. J. Chem. Phys. *91*, 211–218.

Pastor, R.W., Brooks, B.R., and Szabo, A. (1988). An analysis of the accuracy of Langevin and molecular dynamics algorithms. Mol. Phys. *65*, 1409–1419.

Perutz, M.F., Wilkinson, A.J., Paoli, M., and Dodson, G.G. (1998). The stereochemical mechanism of the cooperative effects in hemoglobin revisited. Annu. Rev. Biophys. Biomol. Struct. *27*, 1–34.

Rader, A.J., Vlad, D.H., and Bahar, I. (2005). Maturation dynamics of bacteriophage HK97 capsid. Structure *13*, 413–421.

Ryckaert, J.P., Cicotti, G., and Berendsen, H.J.C. (1977). Numerical integration of the Cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. J. Comput. Phys. *23*, 327–341.

Shimada, J., and Shakhnovich, E.I. (2002). The ensemble folding kinetics of protein G from an all-atom Monte Carlo simulation. Proc. Natl. Acad. Sci. USA *99*, 11175–11180.

Shoemaker, B.A., Wang, J., and Wolynes, P.G. (1999). Exploring structures in protein folding funnels with free energy functionals: the transition state ensemble. J. Mol. Biol. *287*, 675–694.

Snir, M., Otto, S., Huss-Lederman, S., Walker, D., and Dongarra, J. (1998). MPI: The Complete Reference, Second Edition (Boston: MIT Press).

Snow, C.D., Nguyen, H., Pande, V.S., and Gruebele, M. (2002). Absolute comparison of simulated and experimental protein-folding dynamics. Nature *420*, 102–106.

Stock, D., Gibbons, C., Arechaga, I., Leslie, A.G.W., and Walker, J.E. (2000). The rotary mechanism of ATP synthase. Curr. Opin. Struct. Biol. *10*, 672–679.

Tama, F., and Brooks, C.L., III. (2005). Diversity and identity of mechanical properties of icosahedral viral capsids studied with elastic network normal mode analysis. J. Mol. Biol. *345*, 299–314.

Tama, F., Valle, M., Frank, J., and Brooks, C.L., III. (2003). Dynamic reorganization of the functionally active ribosome explored by normal mode analysis and cryo-electron microscopy. Proc. Natl. Acad. Sci. USA *100*, 9319–9323.

Tirion, M.M. (1996). Large amplitude elastic motions in proteins from a single-parameter, atomic analysis. Phys. Rev. Lett. *77*, 1905–1908.

Tozzini, V. (2005). Coarse-grained models for proteins. Curr. Opin. Struct. Biol. *15*, 144–150.

Ueda, Y., Taketomi, H., and Gō, N. (1975). Studies on protein folding, unfolding and fluctuations by computer simulation. I. The effects of specific amino acid sequence represented by specific inter-unit interactions. Int. J. Pept. Protein Res. *7*, 445–459.

Vendruscolo, M., and Domany, E. (1998). Pairwise contact potentials are unsuitable for protein folding. J. Chem. Phys. *109*, 11101–11108.

Wang, Y.M., Rader, A.J., Bahar, I., and Jernigan, R.L. (2004). Global ribosome motions revealed with elastic network model. J. Struct. Biol. *147*, 302–314.

Yang, S., Cho, S.S., Levy, Y., Cheung, M.S., Levine, H., Wolynes, P.G., and Onuchic, J.N. (2004). Domain swapping is a consequence of minimal frustration. Proc. Natl. Acad. Sci. USA *101*, 13786–13791.

Zheng, W., and Brooks, B. (2005). Identification of dynamical correlations within the myosin motor domain by the normal mode analysis of an elastic network model. J. Mol. Biol. *346*, 745–759.

Zheng, W., and Doniach, S. (2003). A comparative study of motor-protein motions by using a simple elastic-network model. Proc. Natl. Acad. Sci. USA *100*, 13253–13258.

Zheng, W., Brooks, B.R., Doniach, S., and Thirumalai, D. (2005). Network of dynamically important residues in the open/closed transition in polymerases is strongly conserved. Structure *13*, 565–577.